International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.8 (2014), pp.83-92
http://dx.doi.org/10.14257/ijmue.2014.9.8.08

Multi-Target Tracking and Behavior Analysis Method for Video Surveillance Applications

Jie Su[a,b] , Gui-sheng Yin[b], Chen Hailong[a] and Luo Zhiyong[a]

[a]*Harbin University of Science and Technology, Harbin, China*
[b]*Harbin Engineering University, Harbin, China*
sujie0366@sina.com

Abstract

In order to obtain a satisfactory performance of visual tracking and video surveillance in complex dynamic scenes without the supervision of qualified workers, an efficient visual detection and tracking method is proposed, which can realize target counting and surveillance, behavior analysis and abnormal detection. Multi-targets tracking method based on novel Bayesian tracking model can manage multimodal distributions without explicitly computing the association between tracked targets and detections. The proposed algorithm is compared with recent works, which shows that it is robust to erroneous, distorted and missing detections and it can be applied in security and management of access points.

Keywords: *Multiple target tracking, Bayesian tracking model, Particle filtering; Behavior analysis, Abnormal detecting*

1. Introduction

Automatic visual target counting and video surveillance have important applications for security fields, such as security and management of access points. However, in order to obtain a satisfactory performance these technologies need professional and expensive hardware, complex installations and setups, and the supervision of qualified workers. In order to minimize the cost, some works have focused on developing detection and tracking algorithms. The typical and simple method is based on background subtraction techniques [1]. However, it has the difficulties of false alarms, noisy detections, missing detections, and split and merged detections. It will loss the information caused by noise in images, complex target motion, non-rigid or articulated nature of targets, partial and full target occlusions, complex target shapes, scene illumination changes, and real-time processing requirements. There are many existing approaches deal with occlusions: Algorithm [2] recognizes the start and end of occlusion frames through merging or splitting dynamic targets, and applies different template search approaches for data association between detected blobs and targets. Riemannian manifold trackers with a single camera were applied in paper[3] where dynamic learning is applied to mitigate the tracking drift. All these methods can handle occlusions to some extent, but become less feasible when targets undergo long-term full occlusions. Algorithm [4] predicts the extent possible by the use of MSPF for each vehicle in the next frame, uses different detection strategies for simple or multiple targets to avoid a global search and improve the tracking speed; by constructing the importance density function based on the latest observations, the algorithm can achieve an accurate and robust tracking in the part of the block and cross-vehicle. Paper[5] proposes an efficient BPR-L1 tracker with minimum error bound and occlusion detection, which employ a two stage sample probability scheme, where most samples with small probabilities from first stage are filtered out without solving the computational expensive. However, these two approaches are more complexity.

ISSN: 1975-0080 IJMUE

Paper [6] proposed the method of reducing interference to the background by using the scheme of correction histogram weighting. It tried to improve the efficiency and robustness by reducing the iterations of mean-shift method and the sensitivity to the initializing of targets. Paper [7] proposed the method of introducing partial background information and the priori information of joint targets when describing targets. After obtaining the posterior probability values of each point within the search area, the Mean-shift method are used to realize relocations and the steady tracking to the target under background clutter, illumination changes and partial occlusion. Professor William Freeman of MIT proposed Bayesian model based on visual perception, which provides a good guidance for my project. Algorithm [8] uses prior information about the geometry of the scene, such as the floor position and the camera calibration to restrict the data association and tracking problems. However, this approach makes more complex the system installation and setting, since it is necessary to compute the camera calibration and estimate the 3D plane of the floor, which in turn depends on the camera location.

Motivated by the above issues, a multi-target tracking and behavior analysis method for video surveillance applications is proposed. The detection stage is based on a parametric background subtraction technique that detects the moving regions in the input video flow. A post processing stage refines the detection by estimating moving regions of the moving targets. The tracking stage uses a Bayesian model to simulate the target trajectories. For this purpose, a particle filtering technique is used to predict a set of hypotheses that represent the most probable target locations. These hypotheses are verified using a novel likelihood function that evaluates each hypothetical target configuration with the set of available detections without to explicit compute their data association. Thus, a considerable saving in computational cost is achieved. Moving targets statistical method is proposed to realize target counting. Behavior analysis method is proposed to realize abnormal detecting.

This paper is organized as follows. Section II gives the Moving targets detect and statistical method. In this section, moving targets were initially detected using background subtraction techniques and it defined the target area and background area. Curve evaluation topology method was used to accurately define the target templates. Target statistical method was used to compute moving targets and renew target state. Section III describes the multiple targets tracking model. It showed Particle filter tracking model, feature select method and target pre-location tactics. Section IV introduces the application of Security surveillance and abnormal action analysis. Section V shows the results of experiments. Conclusions are given in section VI.

2. Moving Targets Detect and Statistical Method

2.1. Building Moving Targets Templates

If multiple targets are initially discrete independent entities, the areas of targets can be expressed as a number of areas without overlap regions. Through combining local edge information with the overall uniform regional information, we can locate the boundary curve and accurately identify multi-targets in complex environment while reduce blur edges, noise. During the process of tracking targets template information is used to realize quickly identifying and tracking targets.

Background subtraction techniques can help us get the initial regions of targets and background. Suppose the image sequence include multiple targets and initial regions of targets are not overlapped for each others. Background includes non-connectivity regions. The image may be divided into a plurality of separate areas, each containing only one target. Then we can get the accurate region of target in each area using Mumford-shah model, which can realize accurately location for contour curve by combining local edge information with overall homogeneous region information and can restrain noise [9].

Curve of C composed by several closed curves separates the image into $m+n$ regions. If target regions fulfill $\phi > 0$, there are m target regions recorded as $C_{i,inside}(i=1,2,\cdots,m)$ and $c_{i,inside}$ is the characteristic value of $C_{i,inside}$. If background regions fulfill $\phi < 0$, there are n background regions recorded as $C_{j,outside}(j=1,2,\cdots,n)$ and $c_{j,outside}$ is the characteristic value of $C_{j,outside}$. $c_{i,inside}$ and $c_{j,outside}$ are brightness values corresponding to their regions. To identifying multiple targets, Mumford-shah model can be improved as following.

$$\begin{aligned} E(C,c_{i,inside},c_{j,outside}) = \sum_{i=1}^{m}\int_{C_{i,inside}} \alpha_i \,|u-c_{i,inside}|^2\, dxdy + \sum_{j=1}^{n}\int_{C_{j,outside}} \beta_j \,|u-c_{j,outside}|^2\, dxdy + \\ + u\cdot length(C) + v\cdot area(C_{inside}) \end{aligned} \tag{1}$$

Region marks are used to mark up target regions and background regions using a template that is corresponding to the image. Target area labeled *m* is defined by the curve $C_{m,inside}$ and it is the initial template of target *m.*

Template needs to be updated to capture the appearance variations of the target during tracking. New templates will be added to the template base if none of the template is similar to the tracking result. Therefore, the tracker is vulnerable to failures when the tracking result with a large occlusion is added to the template set. To prevent an improper addition to the template set, we propose a method to detect the large occlusion in the tracking result before it is added to the template set.

Trivial templates are the solution. Convert the 1D trivial coefficient vector to a 2D trivial image by reversing the way that the target template is vectorized. Each pixel in trivial image is mapped to the pixel in the same location in the template image. Threshold the trivial image and obtain another 2D binary image that we call an occlusion map. The white pixel in the occlusion map indicates that the pixel is occluded and the black pixel indicates no occlusion. Assume that an occluder is large in size and its intensity is different enough to be separated from small random noise. Therefore, the occlusion is a large connected region in the occlusion map. The occlusion detection is then reduced to finding a white area that is large enough to be classified as an occlusion.

If targets are partly shielded, acquiring edge-feature of targets becomes difficult. When the number of targets in marked template is reduced and contour of some target is change, there may be shielded targets. In this case, we need breaking up targets using the following method.

(1) Compare current marked template with the initial marked template and identify the targets being shielded for each other using edge features.

(2) Mark unshielded part of the targets in template according to the initial marked template.

(3) Recover the contour of each target using Hough transformation and expansion method.

(4) Separately re-mark the shielded part of targets using the method of region marks in template.

2.2. Moving Targets Statistical

Background subtraction can completely split motion information and inter-frame difference method can detect differential image. First set the global threshold of frame image difference to detect motion in the scene area. Before tracking targets, we must judge whether there are targets and the number of the targets. Built target marked templates can be used to match moving targets. The identifying algorithm for targets is shown as following.

(1) Carry out initial detection for moving targets using the method of background subtraction and define moving targets feature.
(2) Build the accurate templates for moving rigid targets using Mumford-shah model.
(3) Determine whether the tracking is end. If it is not end, go to the step (4). Otherwise, exist.
(4) Predict the area of moving targets in the next frame by using joint particle filter model, and by matching the templates, we can identify the number of targets and the moving information of the targets.
(5) If the number of the moving targets has changed, identify the newborn targets and disappeared targets. Build new templates for the newborn targets and renew template base. Go to the step (3).

If the number of targets is decreased, there are two situations. On the one hand, there may be some disappeared targets. In this condition, we can find the disappeared targets by matching templates. On the other hand if the shape of some targets changed, there may be shielded targets and we need judge shielded targets.

3. Multiple Targets Tracking Model

3.1. Particle Filter Tracking Model

In this paper, object tracking is interpreted as a sequential state estimation problem. In state estimation, a system state X_t is changing in time t. The main goal is to estimate $X_{\forall t}$ based on the following information. First is the information of the behavior of moving objects. Second is the measurements Z_t which are obtained from sensors at discrete time steps and correlated with X_t. During the process of tracking, the sensors delivering measurements are cameras and the state denotes a object within the view of the cameras.

Particle filter (PF) approach is advantageous over other methods because of its ability to deal with nonlinear target behavior and non-Gaussian measurement noise. In this method, the particles with weights are used to take the place of the calculus of posterior probability distribution. Delivery of particle information among multi-frames can overcome the Sidedness of information of traditional "peak" and this method has strong robustness under partial occlusion environment. The PF works by approximating the underlying probability distribution of the track with a set of weighted samples particles, where the more particles used the more accurate the approximation. Because the importance density function is based on the latest observational information, the algorithm is more correspond with the true state of the posterior probability distribution, which reduces the complexity of global search and the uncertainty of objects.

According to the theory of particle filter, if the target state x_{k-1} is known, its follow-up states x_k can be obtained by computing posterior probability density with the formula 2.

$$p(x_k \mid z_{1:k}) \propto p(z_k \mid x_k) p(x_k \mid z_{1:k-1}) \tag{2}$$

where $p(x_k \mid z_{1:k})$ is expressed with the weighted posterior sampling particle collection shown as the formula 3.

$$p(x_k \mid z_{1:k}) = \sum_{i=1}^{N} \pi_k^{(i)} \delta(x_k - x_k^{(i)}) \tag{3}$$

where N is the number of particle, $x_k^{(i)}$ is the target states obtained according to the state equation and δ is the Kroneck function. $\pi_k^{(i)}$ is the particle weight obtained according to observational data $z_{1:k}^{(i)}$ and $\pi_k^{(i)}$ must fulfills the formula 4.

$$\pi_k^{(i)} \propto \frac{p(z_k \mid x_k^{(i)}) p(x_k^{(i)} \mid x_{k-1}^{(i)})}{q(x_k^{(i)} \mid x_{0:k-1,z_k}^{(i)})}, \ \sum_{i=1}^{N} \pi_k^{(i)} = 1 \tag{4}$$

Note that $q(x_k^{(i)} | x_{0:k-1}^{(i)}, z_k)$ is importance distribution. In the moment k, $\pi_k^{(i)}$ fulfills $\pi_k^{(i)} \propto p(z_k | x_k^{(i)})$ and when $p(x_k^{(i)} | x_{k-1}^{(i)})$ is given, the estimation to state x_k is approximately expressed with $\overline{x_k} \approx \sum_{i=1}^{N} \pi_k^{(i)} x_k^{(i)}$.

The PF based tracker is consisted with six steps: Predict, Compute Histograms, Compute Particle Weights, Normalization, Inference, and Resample. The key steps to realize tracking with particle filter framework are shown as follows:

(1) Initialize particles. Initial particles select edge position coordinates of targets and they are expressed as $\left\{(S_t^{(i)}, w_t^{(i)}) | i = 1, 2, \cdots, N\right\}, \sum_{i=1}^{N} w_t^{(i)} = 1$, where i is the ith particle of time t and S is the position coordinates of targets. Calculate the weights of particles using the characteristics of the Mean shift algorithm which make the particles moving in a direction to increase the maximum density in order to achieve the purpose of implementing filter algorithm using few particles instead of most of the particles, and can calculate the range of targets with a smaller amount of calculation.

(2) Estimate prior distribution $p(x_0)$. The contours of moving targets and their edge-feature sets can be acquired using the given improved Mumford-Shah method. We can define the initial distribution of $p(x_0)$.

(3) Define the prediction model of $p(x_k | x_{k-1})$.

(4) Define $p(z_k | x_k)$ to evaluate $x_k^{(i)}$.

Multi-target tracking can be posed as the problem of estimating the distribution of joint configurations with a single filter, the joint filter. Although powerful due to its generality, this approach will result in expensive computations. The representation size, as well as the cost of computing the update recursion, grows exponentially with the number of targets [10].

A simpler solution would be to represent. Estimate the evolution of the targets independently by instantiating a single tracker for each target. The dynamical model and observation model are assumed to be separable, which leads to separable posteriors under the assumption that the initial distribution is also separable. While this approach scales linearly with the number of targets, it is blind to any kind of interaction.

Now the target set is described with a vector X and its components x^i is the state of target i . Joint distributions of Multi-targets are approximated described as the follows:

$$p(X_t | z_{t-1}) \approx \prod_i p(x_t^i | z_{t-1}), \quad p(X_0) = \prod_i p(x_0^i) \tag{5}$$

Where X_t is the state of multi-targets at the moment t . $p(X_0)$ is the initial distribution of multi-targets.

For reasonably high observation rates, the dynamic component of the propagation process underlying $p(x_t | x_{t-1})$ is dominated by independent one target laws q , a joint factor may be added afterward. The forward model is then defined according to the following formula:

$$p(x_t | x_{t-1}) \approx p(x_t) \prod_i q(x_t^i | x_{t-1}^i) \tag{6}$$

Because we can identify the regions of targets using improved Mumford-Shah method in marked template, their appearances no longer depend on each other. The observation model can be separated in image space, which is shown as the follows:

$$p(z | x) = p(z^0 | x) \prod_i p(z^k | x) \tag{7}$$

This defines an observation model that is robust to partial and complete occlusions and complies with statistical independence of target blobs.

3.2. Feature Selecting and Extracting

Tracking results is largely dependent selection of feature. Classical tracking algorithms choose the color, edge and texture features of the target area to characterize the target. However, during the process of tracking, if the features of target and background are similar, tracking will be fail because the background and target cannot be distinguished. Local background information is used to enhance ability of identifying target in this paper.

Target is portrayed as a point set with weights, which is defined as $D=\{(V_i^j, w_i^j)\}_{i=1}^N$, where V_i^j is the jth feature of point i and w_i^j is the weights of the jth feature and point i.

By obtaining feature attributes of tracking targets in one frame, feature matching can find the position having the maximum similarity with these feature attributes as the target position of this frame in the area to be selected next frame. Combining target template with edge feature can realize accuracy tracking using few particles, and reduce computational complexity.

3.3. Targets Area Prediction and Targets Tracking

According to the theory of particle filter, if the target state x_{t-1} is known, its follow-up states x_t can be estimated using Priori knowledge and posterior probability density model. Similarly, we can estimate the states of x_{t+1}, x_{t+2}, and so on. We can predict the area for each target applying this theory, improve the accuracy of identifying targets, and simplify the computation. Target tracking method with the ability of prediction is shown as the figure 1.

In a pure software implementation, the complexity of Predict, Compute Histograms, Compute Particle Weights, and Normalization is O(N). However, the complexity of Resample can be nonlinear. Complexity of Compute Histograms is also a function of the observation because of the scaling factor. The accuracy of the filter is a strong function of N as well. Domenic forte and Ankur Srivastava carry out an detailed analysis to computational complexity and power [11]. The procedure of particle sample, resample and the complexity analysis are shown in the paper [12].

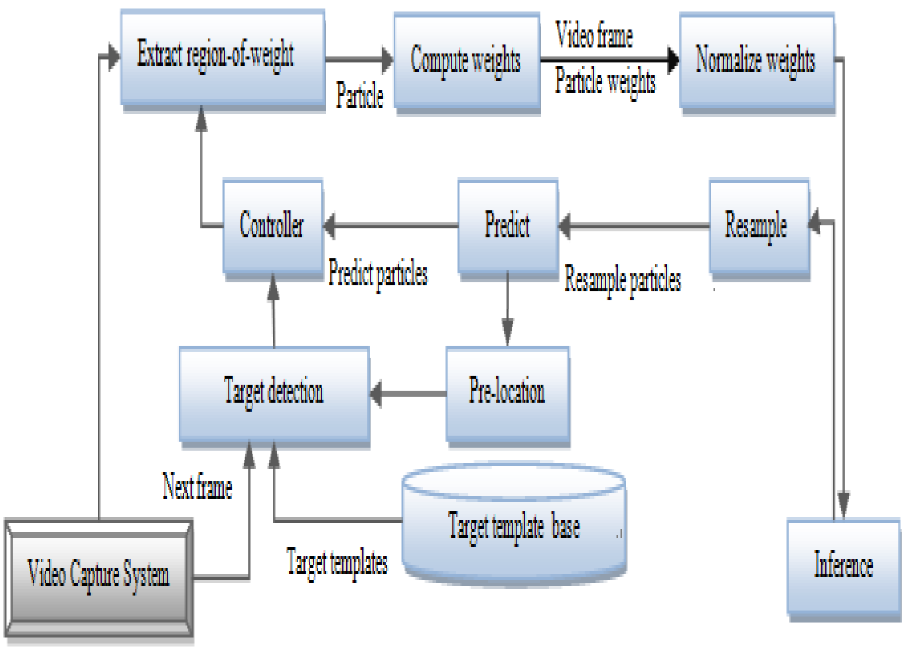


Figure 1. Targets Area Prediction and Targets Track

4. Security surveillance and abnormal action analysis

In monitoring scenarios, requirement of real-time is higher than tracking accuracy, and tracking results are the inputs of high-level algorithm (*e. g.*, behavior understanding and description) in the form of Trajectories.

In observing scene of moving targets identifying abnormal behavior is an important content of high-level decision-making. There are several different criteria that can be used to judge abnormal behavior. For example a behavior of target may be judged abnormal if the target has not be traveling according to the established routes or if the target repeated same behavior. We need to record the relevant parameters in the observation process and timely alert. In no prosecutions video environment, template base of target and behavior predict of the target can help us find abnormal behavior and achieve safety monitoring. The method of behavior analysis is shown as the Figure 2.

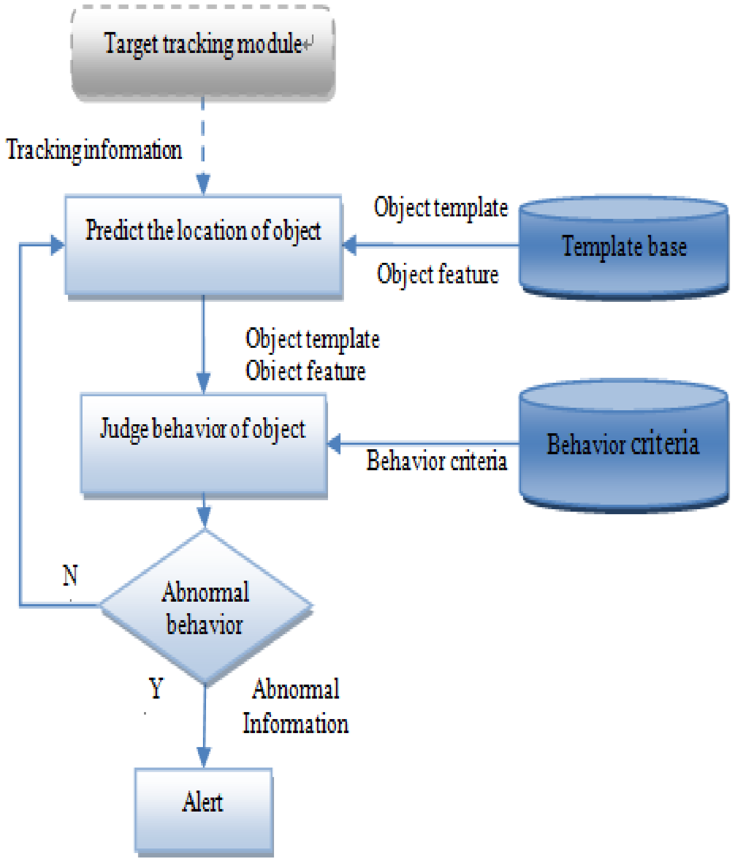


Figure 2. Method of Behavior Analysis

The basic steps are shown as following:

(1) Predict the location of target. Target tracking system will predict the position of the target in the next moment according to the target template and target feature information, which is denoted as the state S_t in time t, and can be acquired according to the S_{t-1}. The state S_{t_n} in time t_n can be deduced according to the information of previous states denoted with $S_{t_n}, S_{t_{n-1}}, \cdots, S_{t_1}, S_{t_0}$.

(2) Judge behavior of target. Remember movement trajectory information and the prediction target trajectory information of target and judge abnormal behavior of target according to the behavior criteria of behavior base.

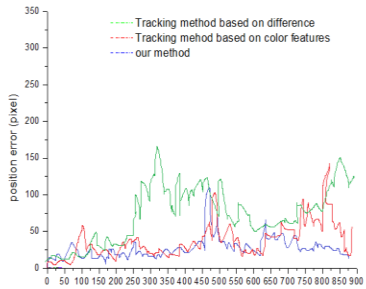
(3) If there is abnormal behavior, there will alert. Else go to the step of (1).

Repeat the above procedure until ending monitoring.

5. Experiments

Experiments are carried out assuming that multi-targets are slowly moving. There are some problems in real time tracking because of the complex computation. In our method, local sample can limit the number of particle and then improve the real-time tracking. The format of the video stream is avi and the videos are denoted as the following video whose number of frames were 600, including multiple random moving targets. The challenging factors in these sequences include occlusion, cluttered background, and mobile multiple targets. The result of experiments is shown with a rectangular box with lower degree of freedom to represent the target state $B_k = \{u_x, u_y, W, H\}$, where u_x and u_y is the center, H is the height of the rectangular and W is the width of the rectangular. Because the target is continue moving, we consider the fringes part of target as the background. Testing the performance of tracking algorithms using fail tracking pixel denoted with $F_{fail}(k)$, where $F_{fail}(k)$ is the number of failed tracking pixel in k frame. Respectively use the classical differential tracking algorithm, color based tracking algorithm and our algorithm tracking the same targets in video sequences.

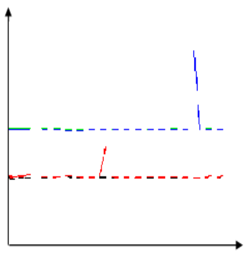
Figure 3 (a) were the results of sequences taken from video sequence frames 32, 128, 201, 264, 332, 376, 438 and 503, which shows that the proposed algorithm is able to identify the contours of different moving targets and can track slowly moving multi-targets accurately. Fail tracking is shown in Figure 3 (b). Figure 4 shows the predicted trajectory and the actual trajectory of moving targets on video. There was a fail estimate on frame 302 and then resulted a false alarm for this estimated result did not comply with the security policy.



(b) Position error of tracking slowly moving multi-targets

Figure 3. Results an Analysis of Track Multi-targets

(a) Predict of tracked target



(b) The points of alarm

Figure 4. Trajectories of Tracked Target

2

6. Conclusions

A multi-target tracking and behavior analysis method applying to video surveillance has been proposed in this paper, which has an ideal result without professional and expensive hardware. This is achieved by the combination of a moving detection algorithm that can handle split and merged detections, and the use of a novel Bayesian tracking model that can handle multimodal distributions, false detections, and missing detections. However, there is a limitation. The algorithm cannot judge and track accurately when the shape of tracking objects changes frequently. Future research works will be focused on it.

Acknowledgements

This research is supported by Foundation of Heilongjiang Educational Committee (12521115), the State Scholarship Fund of China Scholarship Council, the State Scholarship Found of Chinese Scholarship Council (File No.201308930006), the Natural Science Foundation of Heilongjiang Province of China (A201301) and the Heilongjiang Postdoctoral Science Foundation (LBH-Z12082).

References

[1] M. Piccardi, "Background subtraction techniques: a review", IEEE Proc.of International Conference on Systems, Man and Cybernetics, (**2004**), pp. 3099-3104.
[2] G. Chao, S. Jeng and S. Lee, "An improved occlusion handling for appearance-based tracking", in Proc. IEEE Int. Conf. Image Process, (**2011**), pp. 465–468.
[3] Z. H. Khan and I. Y. H. Gu, "Bayesian online learning on Riemannian manifolds using a dual model with applications to video target tracking", ICCV Workshop Inf. Theory Comput. Vis. Pattern Recognit, (**2011**), pp. 1402–1409.
[4] X.-H. Wang, L.-L. Fang and Z.-H. Cong, "Research on Real-time Multi-target Tracking Algorithm Based on MSPF", ACTA AUTOMATICA SINICA, vol. 38, pp. 139-144, (**2012**).
[5] X. Mei, H. Ling, Y. Wu, E. Blasch and L. Bai, "Minimum errorbounded efficient 1 tracker with occlusion detection", in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., (**2011**), pp. 1257–1264.
[6] J. Ning, L. Zhang, D. Zhang and C. Wu, "Robust mean-shift tracking with corrected background-weighted histogram", IET Computer Vision, vol. 1, (**2012**), pp. 62–69.
[7] C. Jun, L.-H. Du, L.-F. Wang and C.-H. Pan, "Local Background-aware Target Tracking", ACTA AUTOMATICA SINICA, vol. 386, Papers (**2012**), pp. 1-11.
[8] K. Yam, W. Siu, N. Law and C. Chan, "Effective bidirectional people flow counting for real time surveillance system, IEEE Proc. International Conference on Consumer Electronics, (**2011**), pp. 863-864.
[9] Y. Y. Guo and X. N. Luo, "Multi-Targets extracted based on single level set", Chinese Journal of Computer, vol. 30, (**2007**) , pp. 120-128.
[10] D. forte, A. Srivastava, "Resource-Aware Architectures for Adaptive Particle Filter Based Visual Target Tracking", ACM Transactions on Design Automation of Electronic Systems, (**2013**) vol. 22, pp. 1-27.
[11] X. Mei, H. Ling, Y. Wu, E. P. Blasch and L. Bai, "Efficient Minimum Error Bounded Particle Resampling L1 Tracker With Occlusion Detection", IEEE TRANSACTIONS ON IMAGE PROCESSING, (**2013**), pp. 2661-2675.
[12] O. Lanz, "Approximate bayesian multibody tracking", IEEE Trans. On Pattern Analysis and Machine Intelligence, 1436-1450(2008).